DOCUMENT RESUME

ED 202 893 TM 810 316

AUTHOR Stecher, Brian
TITLE Constraints and Encouragers to Evaluation Utilization at the School Level.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY National Inst. of Education (ED), Washington, D.C.
PUB DATE [81]
NOTE 31p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Decision Making; Elementary Education; *Interviews; *School Districts
IDENTIFIERS Evaluation Problems; *Evaluation Research; *Evaluation Utilization

ABSTRACT

To obtain a better understanding of evaluation utilization in school-level decision making, principals, special program coordinators, and resource teachers in 22 elementary schools in a large urban school district were interviewed. Some of the more prominent elements in the evaluation process that constrain or encourage evaluation utilization were analyzed. A "constraint" to evaluation utilization was defined as something that a typical administrator would find limits his/her choices or understanding. An "encourager" to evaluation utilization was something that increases an administrator's understanding. Three of the general features that emerged from the data--proximity, competing demands on time, and psychosocial variables--were discussed and their actions as constraints and encouragers to evaluation utilization were analyzed. Sharers, i.e., cooperative decision makers, made the greatest use of evaluation information that arose at the local level and had more direct personal contact. Confronters, i.e., decision makers who were more directive about change, appreciated the clout that came from evaluations generated by higher authority at other levels of the organization. A number of suggestions were made to increase evaluation utilization at the site level. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED202893

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

CONSTRAINTS AND ENCOURAGERS TO EVALUATION UTILIZATION AT THE SCHOOL LEVEL

By

Brian Stecher

Evaluation Use Project

Center for the Study of Evaluation

UCLA

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

M. C. Alkin

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 810 316

The project presented or reported herein was performed pursuant to a grant from the National Institute of Education, Department of Education. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred.

Introduction

To obtain a better understanding of the use of evaluation in school-level decision making we interviewed principals, special program coordinators and resource teachers in 22 elementary schools in a large urban school district. Our conversations focused on decisions that were directly related to events the respondents saw as "significant program occurrences." One of the features that emerged in our analysis of the interviewed data was the differential impact that evaluation had under different circumstances. In some instances school-level decision makers paid careful attention to evaluation when making program related decisions, while in other situations evaluation had almost no impact.

What accounted for these differences? In discussing their actions, our respondents made repeated references to features of the school environment, to the nature of the evaluation activity itself and to interactions between elements in the evaluation process. Upon closer investigation a number of similar, though complex, patterns emerged. We were able to identify constellations of factors that acted on a broad basis either to inhibit the attention paid to evaluation or to enhance its use by decision makers. This paper describes these patterns and analyzes some of the more prominent elements in the evaluation process that constrain or encourage the use of evaluation in decision making at the school level.

This paper will proceed in four stages. First, we will define the terms "constraint" and "encourager" as they emerged in our study. Second, our analysis of three features -- proximity, competing demands on time, and psychosocial variables -- will be presented. Third, we will consider the generalizability of these results based on our experiences in other school

systems. Finally, we will make some recommendations for evaluation practice based on this analysis.

Constraints and Encouragers Defined

Our purpose in undertaking this study was to determine how site-level school administrators used evaluation in actual program decision making. Our respondents described many different kinds of program occurrences, and recounted the thoughts and considerations that affected their actions.

Amid the wide diversity of actions and reactions, certain features were alluded to repeatedly as important determinants of evaluation use. As a simple example, a number of respondents referred to the timing of evaluation results as an important variable in determining whether the data were considered before making a particular decision. In our interview the respondent mentioned elements of the school environment as well as characteristics of the evaluations themselves that affected the liklihood the information would play a meaningful role in decision making. Despite the diversity of the decisions there were great commonalities among these descriptions.

Certain features were mentioned repeatedly as important in determining an administrator's reaction to and use of evaluation. Some things tended to decrease the role that evaluation played in their decisions; other features made them more attentive to evaluation results. The notion of constraints and encouragers was born of these similarities. We will use these labels to identify constructs drawn from reoccurring patterns of attention or inattention reported by significant numbers of respondents in our study.

We will refer to something as a constraint to evaluation utilization if a typical administrator would find that this feature limited his or her choices or understanding. We will refer to something as an encourager to evaluation utilization if a typical administrator would find that this feature increased his or her choices or understanding.[1]

The notions of constraint and encourager put forth here are relative, not absolute. Not all persons would necessarily act in the same manner given the same circumstances. What one person perceives as an insurmountable obstacle to some course of action might be perceived by another person as merely an inconvenient nuisance. Thus, our use of the terms "constraint" and "encourager" is normative. We will only offer as constraints or encouragers those features which represented a substantial commonality across interviewees.

A final note related to these definitions seems in order before describing our results. By focusing on features that were commonly seen to be constraining or encouraging, we do not mean to underestimate the creativity or individual initiative of school administrators. The respondents in our study were a heterogeneous group, and there is probably an exception for every generalization we will offer. There were administrators in our sample for whom even the most frustrating circumstances were not perceived as constraints. Such especially creative individuals are probably worthy of

---

[1]The question may be raised why we bother to specify both terms (constraint, encourager) when they are apparently opposites, and a single definition might suffice. There are two reasons. The first is data based -- they were viewed as distinct entities by our interviewees. Administrators themselves saw certain features as limiting and others as enhancing. It seemed worthwhile to maintain this distinction in our analysis. The other advantage for creating both labels is that certain situations are easier to describe from one point of view rather than the other. While this is a purely syntactic convenience, we decided to retain it because it was so easily accomplished.

additional study themselves. Whether you characterize them as creative, stubborn, self-centered, dynamic or as troublemakers, they were often constrained by factors that inhibited most of their colleagues. They are the outliers in our study, and like outliers in any data analysis, they should be investigated more carefully in the future.

Proximity

We use the word proximity as a generalization of the notion of distance -- how close or far away one thing is from another. But we use the term to mean more than just a spacial comparison. By proximity we mean the degree to which two things are similar or dissimilar along any number of different dimensions. Of particular interest in this study are the dimensions of time and structure (i.e., form, style, content, etc.). Consider the following comments offered by three respondents in our study:

> "Sometimes the district sends us evaluation forms which don't really meet what our school is doing; we try to devise our own based on what they've given us." (11SP1)

> "Well, for the teachers who really are involved in using the test scores from state and mandated tests its helpful... and we do have a couple of teachers who use that. But mostly our teachers use the tests from XXX (the management system), the math program and from the reading program. They mainly use those to see where their children are and to replace and regroup." (13SP1)

> "It seems to me the district needs to get information to the schools more quickly on issues that affect every single classroom teacher, which means those issues affecting every single child within those classrooms." (20SP1)

In one form or another, these three respondents are all talking about the same thing, the proximity of the information to some decision. The greater the 'distance' between the information and an action the less likely it is that the information will be considered in the decision on how to act. The more effort that is required to translate information into a

useable form, retain it until an appropriate time or strip it of emotional overtones the less likely it is that these transformations will be made. In our research, data related to proximity seemed to be easily categorized into two types: structural proximity and temporal proximity.

Structural Proximity

Structural proximity denotes the degree to which a new element matches the format, content, or style of the existing elements of a system.[2] Research has addressed this question to some extent. The impact of reporting format and complexity have been examined by Alkin et al. (1974) and Glaser and Taylor (1975).

Respondents in our study commented frequently on the form and content of the evaluation information available to them. They reported that the configruation of the information -- whether it was directly usable for teachers' instructional decisions -- affected its utilization.

This was manifest among other ways by comments on standardized testing and the district's new criterion-referred test. A large number of interviewees said that standardized tests were less useful than local within-school tests which were based on the school's instructional program. A second common observation was that, among the required achievement tests, the district's criterion-referenced test (DCRT) had the potential to be much more useful than the CTBS test.

---

[2] A further subdivision is possible. One could differentiate between the form of information -- i.e., its physical arrangement and its content. While evidence for such differences can be found in our data, it is a fairly technical distinction which was not generally made. For the time being we will consider only the general category of structure and not subdivide things further.

"We use the XXX reading program, and (evaluate based on) the movement in terms of the number of steps children achieved during the year...The XXX is much easier for use because there's a daily, even a weekly, evaluation...There are so many variables in a one-shot test like the CTBS, so from the school's point of view the XXX management system...(is) much more useful to us." (13P)

"Why are we putting up with this (standardized test) year after year, when we know there are better things we could be doing with our time? There are other instruments possibly which we could be using to give us the kinds of information we want. That's why we lean more heavily on teacher evaluations and those kind of in-house tests." (11SP2)

"The test scores we utilize have been the ones that are criterion-referenced tests like DCRT." (20SP1)

The important point illustrated by these comments pertains not merely to testing but to how easily the information could be used for instructional decision making. The extent to which new information (particularly from evaluation) corresponds to the format and content of information already used by the classroom teacher influences the degree to which it will be used in decisions.

An example from our study is illustrature. A number of schools have adopted the XYZ management system to coordinate their arithmetic program. Students progress is monitored against the XYZ arithmetic continuum in all classrooms. The continuum includes basic arithmetic skills for grades 1 to 6. Learning tasks are prescribed according to a diagnostic test, and students progress through the skill areas one by one. Periodic testing is used to verify the students' mastery of skills and assign new learning tasks.

In the fall of the year teachers at these schools receive the arithmetic test scores from the annual Title I evaluation. The CTBS test is used in this evaluation, and the teachers receive grade level equivalent scores on each student in the areas of Computation, Concepts and Applications.

It should not be surprising to learn that this information is not very useful. There are a number of reasons for this. One prominent reason is that the CTBS scores have little if any direct relation to the XYZ skill levels. The information that the teacher receives from the evaluation based on the CTBS is different, doesn't fit into her regular pattern of assessment, doesn't have a natural correspondence to the ongoing program, etc. It is dissimilar in many respects from the existing classroom structure and each of these dissimilarities is an obstacle to its use. At least that is what respondents in our sample seem to be saying.

It should be pointed out at this point that we are not criticizing the CTBS test on technical grounds -- validity and reliability are not the current issues of concern. Rather we are simply noting that this test (and others like it) is less likely to be incorporated into teachers' planning and decision making if it differs markedly from the data the teacher is already set to process.

We can think of several factors explaining why structural proximity might enhance the utilization of evaluations. Evaluation information that has structural proximity is preferable because:

1. It is familiar and therefore more credible.
2. It requires less effort to translate into a usable form.
3. It matches other data more closely and thus fits more readily into an ongoing aggregate of evaluative data.

These three factors are affirmed by the comments of our respondents:

> "I think that it (school-level evaluation) is more positive because it's at the grass roots. It's more beneficial; its more meaningful because it takes place where the action is... The initial evaluation in a school is teacher-pupil." (20SP1)

> "They (the results from DCRT) are individual, and they are the skills the child needs. But they come back to us in a form

that is not very usable. In order to get the material in a usable form it takes much of the teacher's time, and she's just trying to survive and doesn't quite have that time." (14SP1)

"I think for the most part that the data that's being collected for on-site programs -- School Improvement, Title I -- is, for the most part, useless from one year to the next... The continuum? -- the district changes them. Title I changes their requirements. So from one year to the next, the only continuum at this school that is the same from three years ago when I was Title I Coordinator until now is the one for (the reading program)." (15SP1)

Temporal Proximity

Concerns about time and timeliness were mentioned frequently by the decision makers in our study. Such comments are not surprising since temporal concerns have been identified as important aspects of utilization for almost as long as researchers have speculated about this issue. For example, timeliness has been identified by many writers as being an important variable in utilization (Alkin, 1975; Cohen, 1977; Mitchell, 1973).

We found two different types of temporal concerns. The first we will call timeliness. By this we mean the correspondence between the receipt time of evaluation and the time at which administrative actions are taken.

The bilingual coordinator at one school emphasized the importance of timeliness when discussing annual achievement tests.

"They aren't useful. I don't see teachers using them. You get them late in the year, when you've already planned your program. You know the children by then so they don't give you any new information." (19SP2)

This comment was echoed repeatedly, and there is little doubt from our data that proper fit between delivery of evaluation information and the action schedule affects utilization.

What is interesting to note is how little attention is paid to coordination of evaluation and decision making.

A second aspect of temporal proximity that we noted in our interview was the degree to which data are available for use within their active time frame.[3] Most data have a limited lifespan, and it serves little useful purpose to base decisions on them past their expiration date. For example, achievement scores only remain timely for instructional decisions for a short period; within a month or two the child has learned new skills and his or her old scores on earlier skills are much less useful to the teacher.

The respondents in our study felt that they were burdened with out-of-date data which was of little use to them. Often they were asked to maintain and pass on test scores long past their useful life.

> "I'm saying that as far as an overall tool, to put a great deal of faith in, I don't think it (test data) is worthwhile. For the short run...within a two-week time limit, it's a great tool to take a look at...for the individual teacher who is working with the class...They know who the children are...(But) the teachers are concerned about, 'Why am I passing this data on? The teacher next year really isn't that concerned with it once they start working with those students.'" (15SP1)

To conclude our discussion of proximity we note that respondents made the identifications we have described in this section clearly and distinctly. Little interpretation nor elaboration was required on our part. There was wide agreement on the importance of structure and time. The conclusion we draw based on our data is that increased proximity would likely encourage utilization. Specific recommendation for evaluation in light of this analysis will be presented at the conclusion of this paper.

---

[3] Time can act as a constraint in another way. The pressures and demands of other activities can reduce the available time for consideration of data. We do not consider this as an element of temporal proximity, rather we will discuss it below under the heading of "Competing Demands on Time."

Competing Demands on Time

We use this label to refer to the constraining presence of other job related demands upon decision makers' time. When we asked our respondents how carefully they studied the results of different evaluations, they reported that there were just too many other things demanding their attention to focus extensively on evaluation. As one principal confessed:

> "Well, I'll be real candid with you, I get so busy I don't pay as much attention to evaluation material as I should. I get report after report...I try to get the general gist of what the evaluation data is, but I do not spend a lot of time analyzing it, and I probably should...I think it's probably very good data. There are just so many demands on me." (15P)

There is little doubt that the rapid pace and constant pressure of the school environment constrains administrators' willingness and ability to devote large amounts of time to serious review and analysis of data. Most administrators in our sample reported being innundated by bureaucratic tasks and political pressures. There were some exceptions -- individuals who purposefully guarded their role as educational leader of the school and protected their time, allowing themselves the luxury of contemplation and forethought. But, by and large, most of the administrators we talked to were caught up in the hectic business of running the educational facility keeping up with everchanging regulations, attending meetings, maintaining contact with the community, supervising discipline, and much more.

Both administrators and teachers feld these pressures. Indeed, both reported that their jobs were extremely demanding leaving them little uncommitted time. Here is a typical description, with a suggestion for improvement.

> "I'm sure you must be aware of the fact that a teacher's day is really horrendous in terms of the demands on that teacher's

time. (Teachers need free time to think)...Industry has learned this -- I guess we have learned it, too, but the price tag makes it prohibitive. I think if we could run one pupil-free day a month, or if we could have two pupil-free afternoons a month, or if we had an opportunity to meet together and to interact and to dialogue and share ideas and concerns we would see improvement. But the time constraints are such that it's literally impossible." (13P)

If the demands of the job act as a constraint to utilization of evaluation, is there something that can be done about it? Many of our respondents felt there was a solution. Without specific prompting, a number of decision makers concur with the principal just quoted. They believed that improved use of evaluation data was possible. All that was lacking was the time and opportunity to put some effort in the right direction.

Here are two further opinions on this issue:

"(If we are going to do something with evaluation data) Days have to be set aside...if we could have a few days on the side where the teachers at least sit down and break bread together, I think we'd accomplish a lot more...I don't think there is enough time in the school day to have teachers meet and evaluate the school program. I think if we had some clear days ahead we (would) just sit and talk, one to one, so it's a group. Group discussions to me is the best...I think we need a few days without the children available (to) just sit down and talk about programs." (25P)

"Evaluation tells us where we're going and what we need to do. I think it's very important. I feel that personally I would like to do a lot more of it...But our problem here is (enough time for) meetings, and it does require meetings. I don't think that we evaluate enough. I think we need to have more self-evaluation where we do something like the PQR...once every six weeks is the way I would like to do it. But it seems like we have so many things going on at this school that require teachers to be in meetings...So it's very hard to get people together, even to get a committee together to work on some of these things. I think it needs a lot of improvement." (13SP2)

This thought, that much could be accomplished with the existing evaluation data if only there were time to sit down leisurely, study it and and make plans, was voiced by many of the respondents in our sample. It

was probably the most clearly defined encourager to emerge from the interviews.

Belief in this proposition was strong enough that a few schools had actually attempted to institutionalize opportunities for reflection and reorganization. One school held an annual off-campus conference just before the start of the new school year. They selected a comfortable site (neutral turf, as it were) where the staff could get together without the regular pressures of school to review the accomplishments of the previous year including student test scores and discuss educational activities for the year to come. Another school set aside its last staff meeting for "reflection and projection" during which time the teachers could take a more open and creative look at the school program and the data available from the year just concluded.

Unfortunately, the two instances cited above appear to be exceptional. Not all schools are taking action to fill the need for systematic review and planning time. However, given the existing limitations of budgets and calendars, it is not an easy action to take. In fact, the off-campus conference cited above has been reduced from two days to one this year, and it will be held <u>on</u> campus as well. The school's current budget was just too limited to afford the expense of the previous arrangements.

<u>Psychosocial Variables</u>

The final set of variables that emerged from our data is somewhat more difficult to analyze. Lumped together in this category are psychological variables such as attitudes, feelings and beliefs, and sociological variables such as hierarchical relationships, organizational styles and roles. While structural proximity, temporal proximity and competing demands on time are neat, well-defined constructs with few affective, interpersonal

complexities the psychosocial features of our data are lush with interpersonal and in rrelational complexity. Decision makers reported strong, even intense feelings about certain evaluative processes and information they received, and it was clear that these feelings had an impact on administrative action and on the use of evaluation.

We are not alone in suggesting that evaluation has a strong psychosocial component. Patton (1975) acknowledged this fact when he identified the "personal factor" as the single most important determinant of evaluation utilization. Moreover, feelings and attitudes play an important role in the analytic framework developed in our earlier CSE research. Alkin et al. (1979) included feelings and attitudes explicitly as aspects of the User Orientation dimension, the Evaluator Approach dimension, and the Evaluator Credibility dimensions.

What seems striking about the results of this study is the prominence our respondents gave to the psychosocial components. They were not secondary considerations added to provide additional insight into a respondent's analytic remark. Rather, the psychosocial descriptions were often the principal reaction to a query from the interviewer.

Before we begin this analysis, however, two qualifications are in order. The first concerns the point of entry of this study and one possible reason for the prominence given many of the psychosocial aspects of evaluative interactions. The second relates to the specificity of the respondents' reactions; they spoke more of specific evaluation types rather than evaluation in general.

Point of Entry

While we will give serious attention to the psychosocial dimensions, we nonetheless recognize that their prominence as factors in this study may

be partly an artifact of our point of entry. Specifically, the importance afforded the affective reactions may be due to the level in the organizational structure at which we were making our inquiry. By talking with site-level decision makers, we were focusing on the individuals who actually carried out the instructional program. Most had been classroom teachers themselves, and all were well sensitized to the concerns of teachers. They identified personally with the school, with the teachers, and with the programs that were being carried out.

Our respondents were not policy makers, planners or higher level administrators, who can take a dispassionate view of test results and PQR reports. Rather such data reflected directly on the skills and abilities of the people we interviewed or those to whom they felt closely allied. In short, even though we were asking about program evaluation our inquiries were more easily perceived as personally directed. The data discussed -- mandated tests, pupil assessment, PQR's, etc. -- reflected directly on the abilities of the administrators we interviewed. And, to the extent that evaluations were "close to home," respondents felt that they represented judgments of their personal and professional competence. Some of our respondents were intimidated, often defensive about evaluation. The following report of a school's own on-going monitoring and evaluation committee echoes these concerns.

> "Its (the committee's) job to monitor, review and facilitate change in the program as needed. We've had it for a number of years -- on paper. It is not something that functions with ease. It's a struggle. They don't want to evaluate one another. They don't see it as evaluating the programs, they see it as evaluating one another, and it's a very difficult process...They don't mind checking the evaluation sheets; they don't mind talking about test scores and 'what we're gonna do about.' But, as far as going into rooms and looking at the program in action, nobody wants any part of it. (29SP1)

We believe such remarks and similar expressions of affective sentiments are a valid representation of the concerns of this group of educational decision makers, people who are personally involved in education at the school level. Our intention here merely is to point out that they mey not be representative of decision makers at other levels of the educational delivery system.

Specificity

The final comment we want to make before exploring the various psychosocial dimensions deals with the specificity of interviewee reactions. Simply stated, we found that affective feelings towards evaluation were situation specific. Respondents' views about evaluation generally were of far less importance than their views about specific types. It is not evaluation per se that evoked strong positive or negative feelings, it was rather the PQR or standardized testing or the district E & T consultant. This distinction takes on importance because of the typical lack of differentiation in the research literature between the impact of different evaluation types.[4] Thus, as we untangle the interactions among the various psychosocial dimensions, we will be attentive of discussing differences between evaluation types whenever possible.

Terminology

We found it useful when analyzing the psychosocial responses of respondents in this study to refer back to the framework previously developed by

---

[4]For example, the David (1978) study, while focussing on Title I evaluations, primarily examined the impact of Title I standardized testing data. Alkin et al. (1974) noted the greater impact of formative (as opposed to summative) evaluation. Alkin et al. (1979) examined a variety of evaluation types within their case studies, but were reticent to make generalizations about types because of an insufficient data set.

Alkin et al. (1979). The responses and comments of our subject clustered around four of the dimensions identified in that study. Modifying that terminology slightly we have called these four areas: Evaluation Credibility[5], Origanizational Context, Evaluation Approach and Orientation of the User.

Evaluation credibility represents the degree to which the respondents believe in the results of the evaluation. It is derived from their perceptions of the knowledge and expertise of the personnel who conduct the evaluation, tne use of appropriate unbiased procedures, etc.

The organizational context includes the site level organizational structure as well as the interrelationships between state, district, and site level personnel[6].

The evaluation approach refers to the manner in which the evaluation is conducted and the style and role adopted by the evaluators. Typical of these concerns would be the formal evaluation system or model chat was used as well as the degree of familiarity and personal immediacy of the evaluation process that was undertaken.

The orientation of the users refers to the attitude and expectations of the decision makers themselves.

Comments we obtained from site level administrators shed some light

---

[5]We use the label evaluation credibility, not evaluator credibility as used by Alkin et al. (1979) to indicate that each of the different evaluation types, those that represented the judgment of an identifiable evaluator as well as those that were merely reports of impersonal test data, may differ along the credibility dimension.

[6]Alkin et al. make a distinction between organizational factors within the district and those relationships with institutions' personnel outside the district (They refer to the latter 25 extra-organizational factors.). This distinction is not particularly meaningful in terms of our data, and we will ignore it in this analysis.

on the influence that these variables have on utilization, but they do not yield any definitive understandings. The interrelationships in the psychosocial domain are enormously complex, and we consider this analysis to be only one step in a comprehensive understanding of these factors.

While there were many types of evaluation being carried out in the schools, including locally constructed tests, district developed criterion referenced tests, informal observations, standardized norm referenced tests, formal needs assessments, parent advisory committee program reviews, tests that were part of the existing instructional system, etc., comments relating to variables in the psychosocial dimensions clustered around three specific types. These were the Program Quality Review -- PQR -- (a state-mandated, external, team review process, lasting two to three days), mandated standardized testing (including both norm referenced and criterion referenced tests) and informal, local evaluation activities (such as first-hand observations, informal surveys, shared discussions, etc.). Our discussion of psychosocial variables will refer primarily to these three evaluation activities.[7]

Evaluation Credibility

The question of credibility was alluded to most frequently when the respondent felt it was lacking and this lack acted as a constraint to utilization. Credibility was an issue primarily in discussions of the PQR process and was used to explain the reasons that the PQR was not useful. Two different aspects of credibility were alluded to by our respondents -- the expertise of the evaluators and the procedures that were used to conduct the evaluation.

---

[7]The analysis that follows will be presented in terms of the specific evaluation types discussed by the respondents. However, we believe that the principles which emerged have wider applicability.

Our respondents commented on the knowledge and expertise of the team members who were conducting the review.

> "I don't feel that they are of much value. For one thing the ones we have had were not knowledgeable enough about what the individual school plans are...and the teachers get very defensive..." (27SP2)

> "To make them more valuable to people (at the school) they should be people really knowledgeable about different areas and have a background of knowing what's happening in other places." (19SP1)

> "I find (the state teams) not knowledgeable of inner city programs. Last time we had one from M. and one from another small city. They came in with a very negative attitude..." (11SP1)

> "They're not experts. We're more experts. They have so many hats to wear out there. And they say, 'This week you going to be evaluating a school.'" ( )

These quotes suggest rather clearly that the experience and skill of the evaluator affects the impact of the evaluation.

The second area of concern dealt with the procedures that were employed to conduct the evaluation. When the recipients of the evaluation question the process through which the results were derived it reduces the likelihood that they will incorporate the results into their decisions. Our respondents seemed well aware, in an informal manner, of the ways in which the evaluation process was unreliable or invalid. They recognized when it was unsystematic, when it was not thorough and when the parties did not maintain impartiality.

> "I don't think in that short period of time (when the state team is there) that the State really can adequately evaluate what goes on in a classroom or in a school." (20SP1)

> "And I don't think you can evaluate a school program in one day spending two minutes in a room checking to see if pupil profiles are all done...I mean that's not what it is really about...I found it (the PQR) to be very negative and not telling me much that I didn't know already, and lots of times they're not seeing really important things in our program." (29SP1)

"If you get a very sharp team in, they could probably tell you where you've got a lot of things wrong and where things are right...you don't necessarily get a very sharp team in. The people are human as anybody else and they're not always as able to define everything within the short period of time that they're here." (24SP2)

"Our experience with state MAR team (four years ago) -- everybody's got their own 'bag.' My bag might be learning centers and your bag might be bilingual education, so that when you come to my school there are certain things that you're gonna look for, and you're not gonna see other things. Because you're a human, partial, person -- the teams have functioned as human, partial people. Ideally they should be impartial." (29SP1)

Though most of these negative comments were directed to the PQR, people also questioned the validity and practicality of standardized tests. Here are two comments about standardized testing that illustrate this feeling:

"They're worthless. Those things are part of the mindless tasks of education. Somebody wants them, I don't know who wants them. They're not relevant to our evaluation." (19P)

"...if these children were showing 10 month's growth on a CTBS score they should be showing similar kind of development on the Developmental Reading program and they aren't..." (15SP1)

Before concluding this discussion it is interesting to look at the other side of the coin and ask, what sort of features might increase credibility? We examined our data to see if it shed any light on the question of what makes an evaluation more believable.

Positive comments were reserved for evaluations that were informal, local, more personal -- evaluations they carried out themselves or ones that were carried out by their close associates on the staff. There seemed to be trust and acceptance of this type of informal information. The comments below suggest that the staff itself was deemed to have appropriate knowledge and expertise and that direct, first-hand contact and observation was not frought with the bias, partiality and lack of reliability seen in the PQR.

(Q: Are you referring to formal evaluation?) "No, that would be informal -- powerful -- because people know it and you'll get the same answer from person to person to person." (3P)

A Title I coordinator expressed similar sentiments when lauding the local evaluative activities carried out by the staff in a informal manner:

"I think that it is more positive, because it's at the grass roots. It's more beneficial, it's more meaningful because it takes place where the action is...The initial evaluation in a school...is teacher-pupil." (20SP1)

Overall, there seems to be little doubt that evaluation credibility affects utilization. A lack of credibility due to either perceived lack of expertise and knowledge on the part of the evaluator or to improper procedures will certainly constrain the utilization of the results of the evaluation. High credibility, particularly of the type afforded first hand personal data, seems to function as an encourager to utilization.

Evaluation Approach

Alkin et al. differentiate among seven elements of evaluation approach. Only one distinction emerged clearly in our data -- the importance of personal contact or involvement in the evaluation.

Our respondents frequently mentioned personal interactions between the evaluator and the users of the evaluation. There was little if any comment about the use of formal evaluation models, the research design or the other structural elements of the evaluation approach category developed by Alkin et al. More to the point were phrases the "personal involvement " and "positive rapport."

Decision makers preferred having a sense of involvement in the evaluative process to being passive recipients of evaluation data.

"Getting the people involved and feeling that they have some say-so -- that each one of them becomes an independent information gatherer and sharer, as opposed to all good coming from above.

I am a cog and react. No, we act. We have some say over our professional destiny." (3P)

The importance of positive rapport and personal contact was mentioned in many different forms. For example respondents decried the coldness and negative attitudes of some PQR teams. One principal described an alternative evaluation process he believed would be more effective. What was its strong feature? The evaluator would spend enough time at the school to become one with the staff and gain greater personal understanding and rapport.

"I almost wish that someday we would reach a point where we would hire someone from a university and let them constantly look over things and evaluate and help us in a positive way. But these visits from the state...When you know that's occurring, the staff is not functioning in a normal fashion. They're still wondering what is that rating? Question: You mentioned an evaluator from the university being here more often -- what would be the advantage of that? I think they'd be like part of the staff psychologically. You'd feel they're one of us. You'd feel they're here to work with us." (01P)

All in all, our data suggest that lack of user involvement, and personal contact will all act as a constraint to evaluation utilization.

Orientation of the User and Organizational Context

One of the most interesting patterns that emerged from our interviews was the interaction between the orientation of the users and the organizational context as predictors of evaluation utilization. Two very different orientations toward decision making were described by the administrators in our sample, and the optimum configuration of evaluation elements differed depending on the decision making role and attitude toward evaluation adopted by the administrator. Simply put, the administrator who was oriented toward "shared decision making" or "cooperative governance" paid more attention to evaluations that were local, personal, and included the views of teachers and staff. On the other hand, the administrator who saw change and program

improvement arising through individual leadership and direct action rather than through gradual shared improvement, took advantage of the powerful impact of evaluations that were external, somewhat foreign and carried with them the aura of higher authority.

We refer to the first group (who compared the majority of the respondent in our sample) as the "sharers." They described a cooperative decision making strategy and their orientation toward evaluation reflected this ideal. Most of their positive comments were reserved for informal, local evaluation.

The sharers praised local evaluation efforts and decried formal standardized testing and other external evaluations like the PQR. Their orientation stressed cooperation for program improvement. Three comments typical of this group were:

> "Most formal information we receive from the established agencies either within the district or without the district is of relatively little use to us. We get extensive reports from R & E from Federal government, from Sacramento -- I'm being negative at the moment -- printouts of profiles, percentiles, grade levels. It is of almost no use. It is a waste of time as far as we are concerned...It may be of value to others... The information that comes from people and agencies that purport to serve us are about 99% ineffective. The information that we act upon is generally self-generated, individual type, ferreting out, or visits to programs that exist somewhere else." (3P)

> "We do a lot of formal evaluation that really is worthless. It's worthless. What we get through Research and Evaluation is a lot of statistics that really have no meaning for us... You don't do this formally. You make observations. A lot of people are making observations. Teachers come back and they talk to you. We do a lot of talking to each other, like I'm talking to you." (19P)

> "I feel that the vehicle for change is our conference. The programs are all evaluated there. Whatever we're doing the evaluation is discussion generally with notes. Whatever comes out of the discussion is simply charted. There is no check list or that sort of thing. It's an informal discussion, but that's the evaluation." (3SP2)

Some of their harshest criticism was reserved for the PQR. They felt that the process itself was traumatic and disruptive. (Note: In an earlier incarnation the PQR was called the MAR -- Monitor and Review. Some people still use this old jargon.)

> "I think we had heard scare tales from a lot of people what this MAR team was going to do. They were going to come and 'mar' us in the sense that we'd be 'marred' after they left. (11SP2)

> "They really intimidated the staff. It can be very demoralizing to the staff because they, staff in general -- whether they are or not -- feels as though they're working their tails off." (27SP2)

> "I don't feel anyone should, especially in education, have to go through the feeling of, well, the trauma that is brought up by just watching the PQR. Especially when the trauma is very threatening to that person." (01SP1)

We refer to the other group as the confronters. In contrast to the sharers, the confronters appreciated the impact, the "clout," that formal, external evaluations like standardized testing and the PQR provided. Their orientation is typified by the following comments:

> "We keep on telling the teachers that this needs to be done and that needs to be done. And I don't have that much clout to go in -- no one does except the principal -- to go into different rooms and say 'this is wring,' 'that is wrong.' When a pre-PQR comes, yes, I have a reason to go and visit and say 'this needs to be done, that needs to be done.'" (29SP2)

> "PQR...gives me a mandate, you <u>have</u> to do it...It makes administration easier. It makes change...if you are a change agent, PQR is fantastic." (19P)

> "In the cases where fires are lit under teachers to get their class in order, then the children benefit because you see the teacher putting in extra time...Anxiety, I'd build anxiety... many children are getting short changed. I would build fires. And say '...you either cut it or you don't.'" (3SP1)

> "I think it (the PQR) got some of the teachers who wouldn't have done it otherwise to become more cognizant of what was needed in their program -- to actually look at the record and see what was needed (the written plan). When we worked with the staff we told them, 'You are responsible for seeing that what you are teaching is related to the skills you said you need to teach.'" (13SP1)

The confronters appreciate the formality and distance of the PQR. If old attitudes and well established practices are to be changed one needs to rock the boat somewhat firmly. A local, informal evaluation is not likely to be of much use in this process. On the other hand the seal of high authority that accompanies the PQR of other such evaluation may create larger waves and hence receive greater attention.

To summarize, we found that administrators had two different orientations to decision making and evaluation. Those administrators who took a cooperative orientation to decision making disliked evaluations that arose from other levels of the organization. Administrators who had a directive orientation reported that these same elements of the organizational context acted to enhance evaluation utilization.

We can only speculate about what might cause an administrator to adopt a sharing or a confronting orientation toward a decision. Certainly many factors enter into this determination including personality variables, past experiences, specific training, current staff attitudes, etc. For our purposes it is sufficient to note that these two orientations seemed to exist in our sample to an identifiable extent, and that they respond to very different types of evaluation, particularly in the area of evaluation context. In fact, characteristic that act as constraints to sharers, such as the formality and intimidation of evaluations conducted by the state Department of Education, are perceived as encouragers by the confronters.

There are still many unanswered questions. For example, while our respondents seemed to adopt a consistent sharing or confronting attitude toward both the decisions we discussed in our interviews, it seems more reasonable to believe that this distinction is decision-specific. An administrator might approach one decision as a sharer and another as a

confronter, depending on the nature of the particular situation and the particular individual's ability to adapt his or her style. While our data shed some light on the existence of these distinct styles they do not contain enough information to carry the analysis any further. That will have to await further investigation.

Generalizability

As we noted at the outset, our sample was drawn from a single urban school district, and our focus was confirmed to the perspective of the site level evaluation user. It is important to consider how the results of our study and the recommendations we have made apply in a broader context. How do practical experiences in other districts and at other levels of the educational system affect our analysis and the conclusion we have drawn?

Summary and Recommendations

Three of the general features that emerged from our data -- proximity, competing demands on time, and psychosocial variables -- have been discussed and their action as constraints and encouragers to the utilization of evaluation has been analyzed. Now we will consider some of the practical implications of this analysis for evaluators.

The data suggest that information which is different in form and content from the school's instructional program (structural proximity) is less likely to be used in decision making. Temporal proximity acts in a similar manner. Evaluation is less likely to have impact if it comes at a time when decisions are no longer being made (timeliness) and if the information it contains is no longer current (active time frame) and thus less relevant to the decision.

The second general constraining feature that emerged from our data was competing demands on time. Respondents report that there are so many demands on their time that only a minimum amount of attention can be paid to evaluation. It would seem that more time needs to be provided for review of evaluation data and systematic planning. Certain caution is in order, since the data also suggest that the mere existence of pupil-free afternoons or other open blocks of time will not insure greater attention to evaluation. There are innumerable <u>other</u> demands competing for this free time. To increase utilization the time should be earmarked in some manner specifically for

the purpose of analyzing and acting upon evaluation.

The third feature consisted of psychosocial variables including credibility, approach, context and orientation. We found that the credibility of the evaluation had an impact on utilization. Regardless of the type of evaluation being discussed, if the procedures are not appropriate or the personnel lack expertise in the eyes of the users, the results are likely to be downplayed. It appears that credibility can be enhanced through increased personal contact with the users in the evaluation.

The personal aspect was also the most important concern in terms of the evaluation approach. Both lack of user involvement and poor rapport between evaluators and clients will limit the utilization of evaluation results.

Finally, we explored the interaction between an administrator's orientation to evaluation and the organizational features of the evaluation system. Sharers -- cooperative decision makers -- made the greatest use of evaluation information that arose at the local level and had more direct personal contact. Confronters, who were more directive about change, appreciated the clout that came from evaluations generated by higher authority at other levels of the organization.

This analysis has definite implications for the evaluator or the school administrator who is interested in increasing evaluation utilization. A number of suggestions for improvement can be made in light of our analysis of the constraining or encouraging potential of the form and content of evaluation,

the active time frame of the data, the credibility of the evaluators, the correspondence between orientation and organizational context, etc. To increase evaluation utilization at the site level the evaluation should be planned so that:

1. The data are collected and reported in a form that is easy to use and corresponds to whatever organizational system is in use in the school.
2. The instruments reflect the same content and internal scope as the instructional program at the school.
3. The data collection and reporting process is coordinated with the school calendar and the important identifiable decision periods.
4. The data are analyzed and reported quickly.
5. Time is set aside for review of the information. In this regard, a first-hand presentation with questions may be much better than a written report.
6. Those conducting the evaluation have the appropriate training and expertise.
7. The evaluative procedures are fair and unbiased.
8. Users are involved in the evaluation as much as possible so they develop a positive rapport with the evaluators or a positive attitude toward the evaluation process.
9. There is a match between the kind of impact desired for the evaluation and the manner in which it is conducted.

Our data suggest that following these recommendations will increase the use of evaluation in decision making at the school level. However, one caveat is in order. The fact that something was identifiable as a constraint does not necessarily mean that removing it will increase utilization. If, for example, all evaluation were suddenly structurally and temporally proximate, there might not be any greater use of the information. Administrators might just point elsewhere to explain the continuing non-use of the data. One can never know with certainty the consequences of suggested changes in the way things are done. It is our firm belief, however, that the recommendations derived from this study will have positive impact on utilization.